



HAL
open science

Scottish Health and Ethnicity Linkage Study (SHELS) of 4.65 million people exploring ethnic variations in disease in Scotland: cohort profile

Raj S Bhopal, Colin M Fischbacher, Christopher Povey, Jim Chalmers, Ganka
Mueller, Markus Steiner, David Brewster

► To cite this version:

Raj S Bhopal, Colin M Fischbacher, Christopher Povey, Jim Chalmers, Ganka Mueller, et al.. Scottish Health and Ethnicity Linkage Study (SHELS) of 4.65 million people exploring ethnic variations in disease in Scotland: cohort profile. *International Journal of Epidemiology*, 2010, 10.1093/ije/DYQ118 . hal-00610722

HAL Id: hal-00610722

<https://hal.science/hal-00610722>

Submitted on 24 Jul 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Scottish Health and Ethnicity Linkage Study (SHELS) of 4.65 million people exploring ethnic variations in disease in Scotland: cohort profile

Journal:	<i>International Journal of Epidemiology</i>
Manuscript ID:	IJE-2009-12-1131.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	16-Apr-2010
Complete List of Authors:	Bhopal, Raj; <none> Fischbacher, Colin; NHS National Services Scotland, Information Services Division (ISD) Povey, Christopher; NHS National Services Scotland, Information Services Division (ISD) Chalmers, Jim; NHS National Services Scotland, Information Services Division (ISD) Mueller, Ganka; General Register Office for Scotland Steiner, Markus; University of Edinburgh, Teviot Place, Public Health Sciences Brewster, David; NHS National Services Scotland, Information Services Division (ISD)
Key Words:	ethnicity, cohort, linkage, Ethnic Group, Outcomes

1
2
3
4
5
6
7 **Scottish Health and Ethnicity Linkage Study (SHELS) of 4.65 million people**
8 **exploring ethnic variations in disease in Scotland: cohort profile**
9

10
11
12 **R Bhopal^c, C Fischbacher^{a,c}, C Povey^a, J Chalmers^{a,c}, G Mueller^b, M Steiner^c,**
13 **H Brown,^c D H Brewster^{a,c}, N Bansal^c.**
14

15
16 Addresses

17 ^aInformation Services Division, NHS National Services Scotland, Gyle Square, 1
18 South Gyle Crescent, Edinburgh EH12 9EB.
19

20 ^bGeneral Register Office for Scotland, Ladywell House, Ladywell Road Edinburgh
21 EH12 7TF.
22

23
24 ^cPublic Health Sciences, Centre for Population Health Sciences, University of
25 Edinburgh, Teviot Place, Edinburgh EH8 9AG
26

27
28
29
30 ***Address for correspondence:** Professor R Bhopal, Edinburgh Ethnicity and Health
31 Research Group, Public Health Sciences, University of Edinburgh, Teviot Place,
32 Edinburgh EH8 9AG. Raj.Bhopal@ed.ac.uk
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

How did the study come about?

Many countries require health services to show that they are meeting the needs of ethnic minority populations. (1) This requires data on health status, health care uptake and outcomes and population denominators. Weaknesses in routine data collection often make such requirements difficult to meet. Routine data sources in Scotland, as in most countries, may not include a patient's ethnicity. In Scotland, the need for such data is driven by both policy and legislation responding to rapidly increasing ethnic diversity. Fair For All (2003), Scotland's policy, provides a strategic approach to improve the health of minority ethnic groups(2). The UK Race Relations (Amendment) Act (2000) placed a duty on public bodies to promote racial equality (3). These mandates are reflected in guidance on ethnic monitoring (4-6).

Appropriate service and research is undermined by the lack of data. Ethnic variations occur in all Scotland's national health priority areas, including coronary heart disease/stroke (7;8), cancer (9-11), maternal and child health (12;13) and mental health (14).

In view of the mismatch between need for, and availability of, data by ethnic group Bhopal proposed a demonstration project to explore retrospective approaches.(15)

The project tested proposals including name search methods, analyses by country of birth, modelling/extrapolation from other nations' datasets, and linkage methods. The demonstration project concluded that census-health records linkage methods - in the context of this project first mooted by Povey - held most promise. To our knowledge attempting matching of a national health dataset to a complete national census using demographic identifiers rather than national identity numbers had not been reported

1
2
3 though health data linkage is well established in the UK and internationally,(16)
4
5 including exploring ethnicity and health. (17;18)
6
7
8
9

10 **Who set it up, how, and why, and how was it funded**

11

12
13
14
15 The study was set up in collaboration by the University of Edinburgh (Bhopal and
16
17 Fischbacher), the Information Services Division (ISD) of NHS National Services
18
19 Scotland, and the General Register Office for Scotland (GROS). ISD maintains
20
21 hospital episode data (the Scottish Morbidity Record) that has already been linked to
22
23 GROS mortality data. A permanently linked database joins together mortality to
24
25 general hospital discharge records (SMR01), maternity and birth records (SMR02),
26
27 psychiatric inpatient records (SMR04), and cancer registrations (SMR06). In addition
28
29 we have accessed child immunisation and health records (SIRS, and CHSP-PS) and
30
31 Mental Welfare Commission records which are not in the ISD 'linked' dataset. We
32
33 have linked the census, which contains ethnic codes, to all these data using methods
34
35 reported in detail elsewhere. (15;19)
36
37
38
39
40
41
42

43 The Registrar General of Scotland judged that under present legislation, namely the
44
45 Census Act 1920 and the Census Confidentiality Act 1991, the linkage described here
46
47 was acceptable if the confidentiality of individual personal information was
48
49 maintained during linkage and analysis. We used both one-way encryption methods
50
51 ("hashing") and organisational procedures already published (19) e.g. GROS required
52
53 that the linkage was performed on a stand-alone computer in a locked 'safe haven'
54
55 within GROS premises. Modifications were made to the computer operating system
56
57 to monitor activity and all peripheral devices were disabled. The room at GROS was
58
59
60

1
2
3 accessed with the agreement of GROS, ISD and the PI (Bhopal). GROS maintains a
4 register of visits. We obtained approval from the Scottish Multi-centre Research
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

accessed with the agreement of GROS, ISD and the PI (Bhopal). GROS maintains a register of visits. We obtained approval from the Scottish Multi-centre Research Ethics Committee (MREC) and the Privacy Advisory Committee (PAC) which advises ISD and GROS on the use of data.

This work was supported by the Commission for Racial Equality, NHS Health Scotland (through The National Resource Centre for Ethnic Minority Health), ISD, the Scottish Executive Health Department and the Chief Medical Officer. An ethical opinion, subsequently published, was given by a Professor of Medical Ethics.(20) For funding see acknowledgements.

The figure, republished from open access reference number 19, (19) illustrates in concept how record linkage was based on the use of subsets of information from three datasets: health care records, which include personal identifiers and clinical information; the Community Health Index (CHI) dataset which contains personal identifiers and the CHI number; and the census file which contains personal identifiers and details of individuals' ethnicity. The CHI dataset lists everyone registered with a General Practitioner or eligible for NHS screening services and forms a unique identifier for NHS use. More than 99% of the Scottish population is listed on the CHI.

Date of birth, surname (using soundex codes to allow for variations in spelling), forename, address and full postcode, which were available in both data sets albeit not always recorded identically, were used to link the census number to the CHI. At this stage, other data fields in the two datasets were disconnected from identifying

1
2
3 variables. CHI and the census unique number were encrypted prior to linkage. A one-
4
5 way cryptographic (“hashing”) algorithm (currently impossible to reverse) was used
6
7 to encrypt the CHI number. The census number was encrypted using an algorithm
8
9 developed by GROS. For the records deemed to be matches, 73.6% were exact
10
11 matches. For the remainder, a probability matching process was performed. Here,
12
13 the rate of false positives is critical. Methods have been developed to identify how
14
15 false positives occur and what kind of strategies a human checker employs to decide
16
17 whether a pair match is “good”. (16) These decision strategies were built into a
18
19 “partitioning” computer algorithm. These “partitions” then allow the allocation of
20
21 effort to the most profitable “partitions” which yield the lowest false positive rates
22
23 and highest true positive rates.
24
25
26
27
28
29
30
31

32 Once the linkage was completed personal identifying variables (such as names,
33
34 address, postcode and dates of birth) were removed leaving a file with an encrypted
35
36 CHI number and its corresponding encrypted census number (look up file). A census
37
38 extract containing ethnic code (and limited other data including age, sex and
39
40 indicators of socio-economic status (see Table 1) and an encrypted census number
41
42 was joined to the above look up file using the encrypted census number. The
43
44 encrypted census numbers were then discarded leaving the ethnicity code, some other
45
46 variables from the census and the encrypted CHI number. The relevant parts of the
47
48 ISD linked database were linked via the encrypted CHI numbers. The encrypted CHI
49
50 was replaced with an unrelated serial number (to keep together the multiple records
51
52 on the same people, known as hospital admission spells), resulting in depersonalised
53
54 clinical health records carrying census ethnicity codes. Using methods previously
55
56 described we estimated an upper limit to the false positive linkage rate of 0.08 %.(19)
57
58
59
60

What does it cover?

For phase 1 (demonstration project) we studied myocardial infarction, since it is one of the most important conditions in Scotland, in terms of both mortality and morbidity, is common in ethnic minority groups, and likely to be recorded in hospital morbidity records.(19;21) For phase 2 we are studying cardiovascular disease, maternal and child health, cancer and mental health – see Table 2. We are currently planning phase 3, studying a wider range of conditions and linkage to primary care records. In theory we can study long-term outcomes for any health problem coded in NHS records (and in the records of any other organisations that are able and willing to provide data for linkage). In practice, our work is heavily governed and controlled by data privacy and other ethical considerations, and practical matters of resource, expertise and interest.

Who is in the sample?

The objective was to link all 5 million or so people resident in Scotland in April 2001, the time of the census, but we set an arbitrary standard of 80% or more in each ethnic group, on the basis of a judgement that potential biases would be too great for linkages less than that. About 94% overall were linked in the phase 1 demonstration project using the methods described below, and 95% in phase 2, with the minimum standard of 80% linkage met for every ethnic group (see Table 3), thus creating a retrospective cohort study of about 4.65 million people. Taking into account the census enumeration rate (96%), this represents about 91% of the population resident in Scotland in April 2001.

1
2
3 **How often have they been followed up and what has been measured? How many**
4 **events related to the health outcomes are there?**
5
6
7

8 The cohort focuses on ethnicity (as reported on the census form completed by the
9 head of household) as the primary exposure variable. Other census information is
10 included, and is listed in table 1. Religion (both current and that of upbringing) and
11 country of birth is added to the study database and used to provide a more
12 comprehensive assessment of the relationship between an individual's health and their
13 ethnic group. Socio-economic status is an important co-variable that could potentially
14 explain some aspects of ethnic variation. Accordingly, a careful analysis of the
15 influence of socio-economic status is being built into the analysis. Census variables
16 relating to employment and educational attainment, both of the individual and the
17 head of household, together with area-based socio-economic indicators, are used to
18 build up a comprehensive picture of socio-economic status. The Scottish Index of
19 Multiple Deprivation (SIMD), based on area of residence, and included in many
20 Scottish health databases, is used. This will enable the complex relationships between
21 health, ethnicity and socio-economic status to be elucidated. In addition, data on self-
22 reported long-standing illness are available from the census. We also have travel
23 times from home to the nearest hospital.
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

The outcomes currently under study are given in table 2. The number of events varies,
obviously, by ethnic group, disease outcome, and number of years of follow-up.

Scotland's non-White ethnic minority population is small at about 2%. Table 3 gives
the number of people linked by ethnic group. For White minority populations e.g.
other British and Irish the population size and hence number of events is large. In
phase 2, we are analysing seven years of follow-up for most events. In phase 3,

1
2
3 currently planned, we anticipate nine years of follow-up. For most analysis, given the
4 degree of ethnic inequality, our analyses are statistically robust. For many endpoints,
5 ethnic inequalities are large, with two-fold plus risks being common. Differences of
6 interest (e.g. a 50% difference) can be detected for most end-points. For example,
7 asthma is a relatively important but rare cause of hospitalisation and Indians
8 comprised 0.3% of the Scottish population in 2001 (compared with Pakistanis who
9 were 1%). There are more than 3500 hospital discharges and deaths for asthma per
10 year. With 8 years of outcomes we accrue about 28,000 admissions. Assuming 2% of
11 them are in non-White populations, and that is conservative given previous studies
12 showing a relative excess of hospitalisation in minority populations,(22) that makes
13 560. If 0.3% of the total hospitalisation/deaths are in Indians, that is about 84
14 admissions. With 84 expected events, we can detect a difference of 44% in the rate,
15 with about 80% power and 0.05 statistical significance. In phase 2, we have already
16 demonstrated that robust statistical analyses is possible within this data set for many
17 of the cardiovascular and maternal/child health outcomes in the table (analyses on
18 mental health and cancer are underway but we cannot comment at this point).

43 **What is attrition like?**

44 The 5% of the Scottish population that is not linked is being characterised to assess
45 potential biases by comparing with the 95% that was linked. It is not possible to
46 compare with the additional 4% of the Scottish population that was not enumerated at
47 the 2001 census. People who migrate, or are hospitalised or die, outside Scotland are
48 poorly characterised in our databases and, to date, we have not developed a sound
49 strategy for estimating the potential biases. We particularly welcome collaborations
50 on this challenge.
51
52
53
54
55
56
57
58
59
60

What has it found? Key findings and publications

Overall, 94% in phase 1 and 95% in phase 2 of the 4.9 million census records were matched to a CHI record with an estimated false positive rate of less than 0.1 %.

Phase 1 data showed that between April 2001 and December 2003 there were 126 first episodes of acute myocardial infarction (AMI) among South Asians and 30,978 among non-South Asians. The incidence rate ratio was 1.45 (95% CI 1.17, 1.78) for South Asian compared to non-South Asian men and 1.80 (95% CI 1.31, 2.48) for South Asian women. After adjustment for age, sex and any previous admission for diabetes the hazard ratio for death following AMI was 0.59 (95% CI 0.43, 0.81), reflecting better survival among South Asians(19). Phase 2 analyses of cardiovascular diseases are underway and will be submitted for publication shortly. They seem to both corroborate and extend phase 1 findings e.g. confirming increased incidence rather than increased case fatality in explaining the high CHD mortality rate in the South Asian population and showing heterogeneity between Indians and Pakistanis. For example, directly standardised rate ratios for first MI (in the period May 2001 to April 2008) in males were highest for Pakistanis, followed by Other South Asians, and Indians. For females, rates were higher for Other South Asians, followed by Pakistani and Indian. Pakistani males had significantly higher rate ratios of first MI than their White Scottish counterparts. After adjustment for diabetes and the Scottish index of social and economic deprivation (SIMD), Pakistani males and females had significantly better survival 28 days after first MI than the White Scottish population. Analyses relating to phase 2 goals on cardiovascular disease, cancer, maternal and child health and mental health will be reported in 2010.

What are the main strengths and weaknesses?

We had to balance individuals' right to data privacy (individual consent was impossible) and the potential benefits to society of producing information derived from potentially sensitive data.(20) We and a professional ethicist (20) judge that our methods strike the balance. The proportion of records successfully linked was only slightly lower than that normally achieved within ISD in internal linkages (typically around 98%), even although we linked data held by two separate agencies. As the non-White ethnic minority population (about 2%) is comparatively small the false positive rate is critical. A rigorous matching methodology appropriate for administrative matching was, therefore, used.

Census forms are completed by the public and processed electronically using optical recognition and keying from images. The success of these processes depends on legibility. The detailed spelling of a name, particularly if it is associated with a minority ethnic group, is more likely to be prone to error when transcribed by a third party onto NHS records, than in the census where the census informant writes it. Such errors could lead to varying accuracy of linkage by ethnic group. Nevertheless, we met our prior stated standard of 80% for every ethnic group, lower in non-White than White populations. People who are not matched are not included in our analyses. At present we do not know about the characteristics of those matched compared with those not matched, but work on this is underway.

The major strength of this analysis is the population coverage with the inclusion of community and hospital deaths as the SMR database contains generally validated and

1
2
3 complete data e.g. over 99% of hospital admissions for AMI in Scotland (A Redpath,
4
5
6 ISD, personal communication, 2006). Possible, as opposed to probable or definite,
7
8 AMIs are coded as chest pain and are excluded from AMI figures, so the AMI figures
9
10 could be underestimates. The quality of data on AMI in the database has been
11
12 validated as reliable [19]. Information about the quality of SMR01, SMR04 and
13
14 SMR06 is available.
15
16
17
18

19
20 Validation has not, however, been done by ethnic group. For example, the diagnosis
21
22 of AMI was based on criteria used by clinicians to make the diagnosis, which would
23
24 be inconsistent over time. This should not, however, affect ethnic groups differently.
25
26 If misclassification of diagnosis is non-differential then the differences would
27
28 probably be underestimated, but if they were differential, differences would be
29
30 exaggerated. We do not have data to assess these options, and the anonymised
31
32 methods precluded validity studies. We do know that South Asian populations in the
33
34 UK, and clinicians looking after them, are sensitised to the high rate of CHD, and
35
36 early admission and treatment, (23;24) concordant with excellent survival,(19) is
37
38 likely.
39
40
41
42
43
44
45

46 Linkage rates were slightly lower for South Asians than non-South Asians. If non-
47
48 linkage occurred at random, this would reduce the power of the study, but not bias
49
50 the results. Factors reducing the power of the study are particularly important for
51
52 small ethnic minority populations and rare outcomes. If those not matched were at
53
54 different risk of hospitalisation or death from those who were matched, which seems
55
56 quite likely, this would bias the results. We intend to explore this in future research.
57
58
59
60

1
2
3 Potential sources of bias in the analyses have not yet been assessed systematically. A
4
5 separate data linkage evaluation exercise is planned in collaboration with NHS
6
7
8 Central Register to provide alternative linkage outcomes for a small sample of records
9
10 that can help assess the accuracy of the main data linkage project. A comparison of
11
12 census characteristics of matched and non-matched individuals is also planned.
13
14
15 Quantifying the potential impact of deaths abroad, to account for the possibility that
16
17 minority ethnic groups may be more likely to move abroad when they are elderly or
18
19 ill with chronic conditions (the so-called “salmon bias”) is proving problematic. This
20
21 bias would improve apparent survival. Such an effect is unlikely to be large, as
22
23 reflected in mortality by country of birth analyses in England and Wales showing that
24
25 mortality is not uniformly low in most ethnic minority groups, but we will scrutinise
26
27 our results to detect such bias (e.g. as reflected in relatively good long-term compared
28
29 to short-term survival).
30
31
32
33
34
35

36 The statistical power is governed by the size of the population and event rate which
37
38 determine number of cases, and the size of the differences between comparison
39
40 groups. At the 2001 Census about 2% of the Scottish population was not White (about
41
42 100,000 people) and most of them were South Asian (Pakistani and Indian) or
43
44 Chinese. These groups will be our main foci for analysis, although where there is
45
46 sufficient statistical power for other ethnic groups we will examine them e.g. of
47
48 African ethnicity. We are also paying attention to White minorities in Scotland e.g.
49
50 the other UK (mainly English and Irish) populations. In our phase 2 analysis and
51
52 interpretation we are examining heterogeneity of broadly defined ethnic groups e.g.
53
54 South Asian. For these non-White ethnic groups we are finding major differences in
55
56 event rates in relation to comparison populations. One limitation is that we are unable
57
58
59
60

1
2
3 to study Eastern Europeans and any other ethnic groups which were not separately
4 identified in 2001. (This kind of work may be a focus for future work based on census
5
6
7
8 2011.)
9

10
11
12 Power calculations (80% power at the 5% level of significance) have shown that
13 differences of a size that is common and of interest (e.g. a 50% increase) can be
14
15 detected for most end-points, for the three largest non-White minority populations,
16
17 namely Pakistanis, Indians and Chinese. Our calculations assume 5 years of follow-up
18
19 data but for most outcomes we have 7-years, and this will increase in future linkages.
20
21
22
23
24
25

26 27 **Can I get hold of the data? Where can I find out more?**

28
29 This project has created a retrospective cohort of about 4.6 million people living in
30
31 Scotland at the time of the 2001 Census. Phase 2 work is focusing on cardiovascular
32
33 disease, cancers, mental health and maternal and child health. Subject to ethical
34
35 approval, future work will have potential to examine health care procedures and
36
37 adjust for other variables that available in the census. Research using the datasets
38
39 requires approval from GROS, ISD, a multicentre research ethics committee, Privacy
40
41 Advisory Committee and the Directors of Public Health. Approvals have been given
42
43 *only* to examine ethnic variations in health and healthcare.
44
45
46
47
48
49

50
51 The researchers appreciate the benefits of collaboration and sharing of data, and are in
52
53 principle open to proposals for relevant collaborative analysis of datasets. There are
54
55 two products of this data linkage project that may be of value to other researchers:
56
57

- 58 • The CHI-census linkage file, containing the encrypted census identifier
59
60 numbers, and encrypted CHI numbers (dataset #1) and

- The matched, anonymised dataset containing census variables and health records (dataset #2).

Use of the data files requires the use of specialist software and the expertise of research team members, not least in obtaining necessary approvals. Collaboration will be pursued if the prime objectives of the project are not impaired or delayed in the four areas under study in phase 2. Collaborators would be primarily responsible for funding the additional research, including the cost of extra work for the core research team. We are committed to seek ways of increasing access to the database.

Conclusion

Our approach, potentially, has international applicability. It demonstrates how the glaring absence of cohort studies reporting by ethnic group in Europe can be overcome.(25) There is considerable potential in linking databases that have previously been considered too sensitive for record linkage or where linkage is restricted by data protection legislation. The methods described here and in more detail in our report and publication (19;21) have the potential to fill the information gap on data by ethnic group. This gap will persist until we have high quality prospective ethnic group coding systems in health-care databases and the inclusion of a valid ethnic code on birth and death registration - both formidable long-term challenges, hitherto unachieved in either Europe or North America [1]. It is noteworthy, however, that prospective ethnic coding will not provide the detailed socio-demographic data available in the census, so will not replace our method.

Authors' Contributions

RB was the originator of the project and the principal applicant for the phase 1 grant with JC and CF as co-applicants. RB and CF are joint PI's for phase 2. All authors are coinvestigators or collaborators for either phase 1 or 2, except NB who is the research fellow in phase 2. CP had the idea of linking the census data to the data held by ISD and he performed most of the linkage work in phase 1 and phase 2 including the development of the linkage methods. JC had the original plan for the use of one-way encryption. GM was the main contributor from GROS, making both practical and intellectual contributions. Helen Brown (HB) was the statistical research associate on phase 2 of the project from August 2007- December 2009. HB, NB, CF, RB, and MS (MS for phase 1) led the analytic plan and CF, HB, NB and MS did the analysis. RB and CF are guarantors for this paper. All authors contributed to the design of the study, and contributed to, read and approved the final version of this manuscript which was prepared by RB.

Acknowledgements, Copyright and Funding

We thank the Scottish Executive for supporting the phase 1 study and the organisation that assisted in managing the grant and the work – the Public Health Institute for Scotland (now part of NHS Health Scotland). Numerous people offered their expertise, guidance and support (too many to list comprehensively). David Orr of GROS was intimately involved in the phase 1 study from an early stage and served the Steering Committee of the project until 2004 when his place was taken by Peter Scrimgeour, Kirsty McLachlan and Ganka Mueller. Joe Fuchs, Ian Maté and Susan

1
2
3 Wallace of GROS made many contributions to the linkage component, and Duncan
4
5 Macniven of GROS gave helpful advice, particularly on dissemination of the findings.
6
7
8 Professor Phil Hanlon acted enthusiastically and promptly to help make this study
9
10 possible when it was first mooted in 2002. Dr Rafik Gardee ensured the study
11
12 received the full backing of the National Resource Centre for Ethnic Minority Health.
13
14 Mr Hector Mackenzie of the Scottish Executive, Dr Mac Armstrong (CMO), and
15
16 Chris Oswald (CRE) also played influential roles in maintaining the directions and
17
18 funding of the work. Kate Macintyre advised on analytic methods and interpretation
19
20 of data, using experience in the SLIDE project, under the direction of Simon
21
22 Capewell. Finally, we thank our employing organisations for their support of our
23
24 work in this project.
25
26
27
28
29
30
31

32 Phase 1 was funded under a special grant from the Scottish Executive and phase 2
33
34 with a project research grant (CZH/4/432). The researchers acted independently of
35
36 the funding body at all stages of the work.
37
38
39
40

41 The Corresponding Author has the right to grant an exclusive licence for publication.
42
43
44

45 Reference List

- 46
47
48 (1) Bhopal RS. Ethnicity, race, and health in multicultural societies; foundations for
49 better epidemiology, public health, and health care. Oxford: Oxford University
50 Press, 2007.
51
52
53 (2) Scottish Executive. Fair for All: Working Together Towards Culturally-
54 Competent Services. Edinburgh: Scottish Executive NHS HDL(2002)51:
55 2002.
56
57 (3) The Home Office. Race Relations (Amendment) Act 2000: New Laws for a
58 Successful Multi-Racial Britain. London: Home Office Communication
59 Directorate, 2001.
60

- 1
2
3 (4) The Duty to Promote Race Equality: A Guide for Public Authorities in Scotland.
4 2002. UK, Commission for Racial Equality (CRE).
5
6
- 7 (5) Practical guide to ethnic monitoring in the NHS and social care. 2005. UK,
8 Department of Health (DOH).
9
- 10 (6) Ethnic Monitoring Toolkit & Communication Guideline. 2006. Edinburgh, ISD
11 Scotland.
12
- 13 (7) Gill PS, Kai J, Bhopal RS, Wild S. Health care needs assessment: black and
14 minority ethnic groups. Available from: [http://hcna.radcliffe-](http://hcna.radcliffe-oxford.com/bemgframe.htm)
15 [oxford.com/bemgframe.htm](http://hcna.radcliffe-oxford.com/bemgframe.htm). In: Rafferty J, editor. Health care assessment.
16 The epidemiologically based needs assessment reviews. Third Series.
17 Abingdon: Radcliffe Medical Press Ltd, 2006.
18
19
- 20 (8) Bhopal R. What is the risk of coronary heart disease in South Asians? A review
21 of UK research. J Public Health Med 2000; 22(3):375-385.
22
23
- 24 (9) Bahl V. Cancer and ethnic minorities--the Department of Health's perspective. Br
25 J Cancer Suppl 1996; 29:S2-10.
26
27
- 28 (10) Bhopal RS, Rankin J. Cancer in minority ethnic populations: priorities from
29 epidemiological data. Br J Cancer Suppl 1996; 29:S22-S32.
30
- 31 (11) Wild SH, Fischbacher CM, Brock A, Griffiths C, Bhopal R. Mortality from all
32 cancers and lung, colorectal, breast and prostate cancer by country of birth in
33 England and Wales, 2001-2003. Br J Cancer 2006; 95(3):424.
34
- 35 (12) Collingwood BA. Investigating variations in infant mortality in England and
36 Wales by mother's country of birth, 1983-2001. Paediatr Perinat Epidemiol
37 2006; 20(2):127-139.
38
39
- 40 (13) Gardosi J, Chang A, Kalyan B, Sahota D, Symonds EM. Customised antenatal
41 growth charts. Lancet 1992; 339(8788):283-287.
42
43
- 44 (14) Sashidharan SC. Inside Outside: Improving Mental Health Services for Black
45 and Minority Ethnic Communities in England. [England]. 2003. Department
46 of Health.
47
- 48 (15) Bhopal R, Fischbacher CM, Steiner M, Chalmers J, Povey C, Jamieson J et al.
49 Ethnicity and health in Scotland: can we fill the information gap? A
50 demonstration project focusing on coronary heart disease and linkage of
51 census and health records.
52 <http://www.chs.med.ed.ac.uk/phs/research/Retrocoding%20final%20report.pdf>
53 [f](http://www.chs.med.ed.ac.uk/phs/research/Retrocoding%20final%20report.pdf) . 7-8-2006.
54
55
- 56 (16) Newcombe H. Handbook of record linkage: methods for health and statistical
57 studies, administration and business. Oxford: Oxford University Press, 1988.
58
59
60

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- (17) Jeffreys M, Stevanovic V, Tobias M, Lewis C, Ellison-Loschmann L, Pearce N et al. Ethnic inequalities in Cancer survival in New Zealand: Linkage study. *Research and Practice* 2005; 95:834-837.
- (18) Harding S, Balarajan R. Mortality of third generation Irish people living in England and Wales: longitudinal study. *BMJ* 2001; 322:466-467.
- (19) Fischbacher CM, Bhopal R, Povey C, Steiner M, Chalmers J, Mueller G et al. Record linked retrospective cohort study of 4.6 million people exploring ethnic variations in disease: myocardial infarction in South Asians. *BMC Public Health* 2007; 7(1):142.
- (20) Boyd KM. Ethnicity and the ethics of data linkage. *BMC Public Health* 2007; 7(1):318.
- (21) Bhopal R, Fischbacher CM, Steiner M, Chalmers J, Povey C, Jamieson J et al. Can we retrospectively analyse Scotland's health databases by ethnic group? A demonstration project focusing on linkage of census and health records. *Social Dimensions of Health Institute, International Conference, 18-20 September 2007 Universities of St Andrews & Dundee. SocialDimensions of Health Institute . 2007.*
- (22) Netuveli G, Hurwitz B, Levy M, Fletcher M, Barnes G, Durham SR et al. Ethnic variations in UK asthma frequency, morbidity, and health-service use: a systematic review and meta-analysis. *Lancet* 2005; 365:312-317.
- (23) Chaturvedi N, Rai H, Ben-Shlomo Y. Lay diagnosis and health -care-seeking behaviour for chest pain in south Asians and Europeans. *Lancet* 1997; 350:1578-83.
- (24) Ben-Shlomo Y, Naqvi H, Baker I. Ethnic differences in healthcare-seeking behaviour and management for acute chest pain: secondary analysis of the MINAP dataset 2002-2003. *Heart* 2008; 94(3):354-359.
- (25) Ranganathan M, Bhopal R. Exclusion and Inclusion of Nonwhite Ethnic Minority Groups in 72 North American and European Cardiovascular Cohort Studies. *PLoS Med* 2006; Vol 3(3):0001-0008, <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1324792>.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 1 Overview of Record Linkage Process

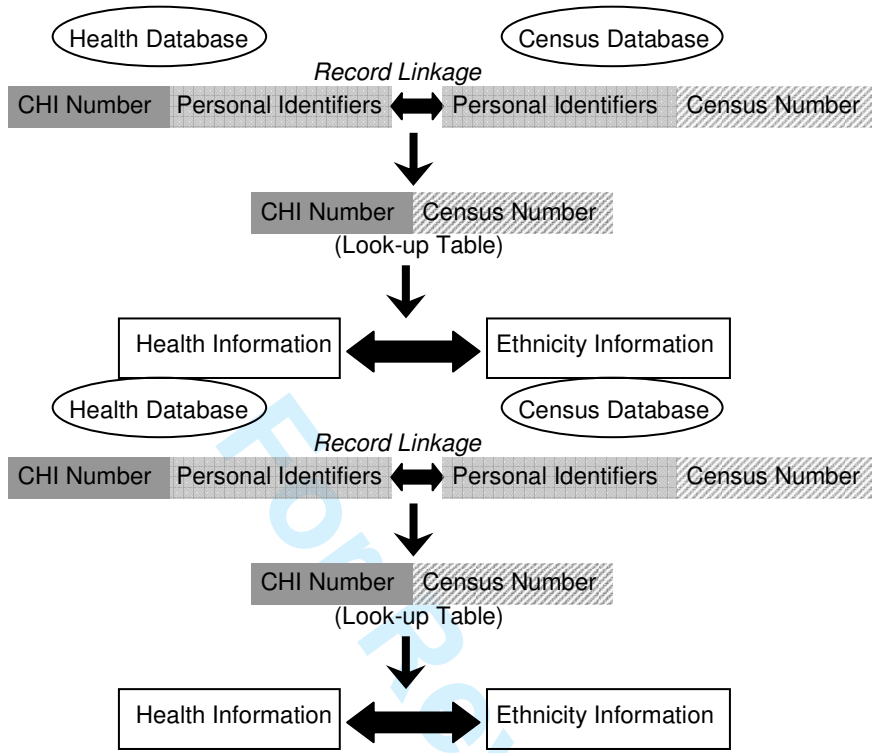


Table 1 Baseline 2001 variables extracted from census: Phase 2

1	
2	
3	
4	
5	
6	Ethnic group
7	
8	Religion, current
9	
10	Religion of upbringing
11	
12	Country of birth
13	
14	Age
15	
16	Sex
17	
18	
19	Long term illness
20	
21	Self assessed health
22	
23	Marital status
24	
25	
26	Labour force status
27	
28	Socioeconomic status
29	
30	Highest qualification
31	
32	Scottish Index of Multiple Deprivation decile
33	
34	Car ownership
35	
36	Housing tenure
37	
38	
39	Household size
40	
41	Number of rooms
42	
43	Urban/rural indicator
44	
45	
46	Health board (Glasgow, Lothian, Tayside, Other)
47	
48	Mobile (temporary) accommodation
49	
50	Self-contained accommodation
51	
52	Central heating
53	
54	Moved within last year
55	
56	
57	Activity last week
58	
59	
60	

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

Table 2 Primary Endpoints for Phase 2 Project

Priority Area	Health outcome/disease/condition	Service Use	Outcome	Datasets
Cardiovascular	Angina	Angiography	Survival following AMI	SMR01,
	Chest pain	Angioplasty	30-day survival following AMI	GROS mortality records
	MI	CABG		
	Heart Failure	Length of hospital stay		
	Stroke			
	All CHD			
Cancer	All CVD			
	Breast Cancer	Uptake of breast cancer screening	Survival following first registration of:	SMR01, SMR06, GROS mortality records, Scottish Breast Screening Programme
	Colorectal Cancer	Stage at presentation	Breast cancer	
	Lung Cancer	Treatment	Colorectal cancer	
Prostate Cancer	Place of death	Lung cancer		
			Prostate cancer	
			(stage and grade to be included in analyses where possible)	
Maternal and Child Health	Age of mother at birth	Mode of delivery	Breastfeeding	SMR01, SMR02 child health surveillance

	Smoking at 1 st visit	Analgesia at delivery	Birthweight	data
	Premature births	Immunisation	Height/weight of infant at health visits	
	Abortion	Child health visits completed		
	Miscarriage			
	Perinatal mortality			
	Hospital Admissions			
	Childhood obesity			
Mental Health	All psychiatric hospital admissions including intentional self-harm	Length of hospital stay Admission type Detention Discharge – type Readmissions	Survival to 2 nd episode of self-harm Survival following attempted suicide Time between admission(s)	SMR01, SMR04 MWC detention data GROS mortality records
	Suicide			

AMI: Acute Myocardial Infarction, CHD: Coronary Heart Disease, CVD: Cerebrovascular Disease, CABG: Coronary Artery Bypass Graft, SMR: Scottish Morbidity Record, MWC: Mental Welfare Commission

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

Table 3 Linkage Rates by Ethnic group

		Number	Percentage
1	White Scottish	4290153	95.3
2	Other White British	357788	93.6
3	White Irish	47173	92.2
4	Other White	74655	87.9
5	Any Mixed Background	12117	91.7
6	Indian	13717	89.9
7	Pakistani	28538	89.8
8	Bangladeshi	1783	88.0
9	Other South Asian	5810	85.1
10	Caribbean	1659	89.5
11	African	4514	86.5
12	Black Scottish or Other Black	1057	89.1
13	Chinese	15115	87.4
14	Other Ethnic Group	8945	86.2

FC Review Only